

Informations-infra-strukturen für Forschung und Wissenschaft: Sicherung, Bereitstellung und Archivierung von Forschungs- daten

Matthias Schulze

This document appeared in

JARA - Forschungszentrum Jülich und RWTH Aachen University (Eds.):

Spezialbibliotheken – Freund und Follower der Wissenschaft

Präsentationen der 33. Arbeits- und Fortbildungstagung der Arbeitsgemeinschaft der
Spezialbibliotheken e.V. – Sektion 5 im Deutschen Bibliotheksverband

Proceedings of the ASpB 2011: Spezialbibliotheken - Freund und Follower der Wis-
senschaft, 09.-11. November 2011, Auditorium der Zentralbibliothek im Forschungszen-
trum Jülich

Zentralbibliothek (ZB)

Forschungszentrum Jülich GmbH, Zentralbibliothek, Verlag, 2011

ISBN:



Informations-infra- strukturen für Forschung und Wissenschaft: Sicherung, Bereitstellung und Archivierung von Forschungsdaten

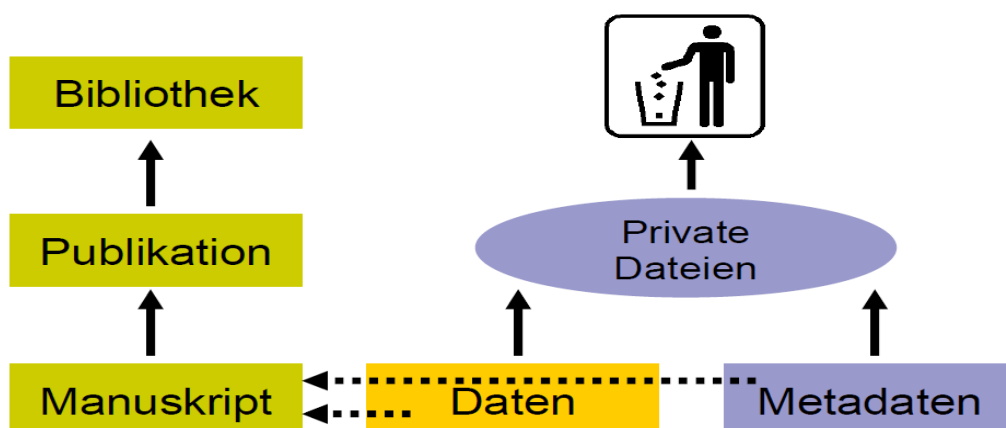
Matthias Schulze
Universitätsbibliothek Stuttgart



http://www.carl-abrc.ca/about/working_groups/pdf/data_mgt_toolkit.pdf

Schulze - ASpB-Tagung, Jülich, 9. November 2011

Umgang mit Daten heute



Vorbemerkung

- „In den allermeisten wissenschaftlichen Vorhaben entstehen **digitale Daten** als Grundlage für neue Erkenntnisse. Ein großer Teil dieser Daten – manche Schätzungen reichen **bis zu 90 %** – **gehen** jedoch in einem relativ kurzen Zeitraum **verloren**. Sie stehen somit nicht mehr einer weiteren Verwendung und **Nachnutzung zur Verfügung.**“



Vorwort von Stefan Winkler-Nees aus „Handbuch Forschungsdatenmanagement“, hrsg. von Stephan Büttner, Hans-Christoph Hobohm, Lars Müller, Bad Honnef 2011 S. 5.

<http://opus.kobv.de/fhpotsdam/volltexte/2011/241/pdf/HandbuchForschungsdatenmanagement.pdf>

Schulze - ASpB-Tagung, Jülich, 9. November 2011



Übersicht

- Einführung
- BW-eLabs
- Ergebnisse – Nachhaltigkeit – Herausforderungen
- Ausblick – Perspektiven

Einführung: Relevanz

➤ DFG-Denkschrift „Sicherung guter wissenschaftlicher Praxis“ (1998):

„Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für **zehn Jahre aufbewahrt werden**.“

➤ Schwerpunktinitiative Digitale Information (2010): **Grundsätze zum Umgang mit Forschungsdaten**

<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/>



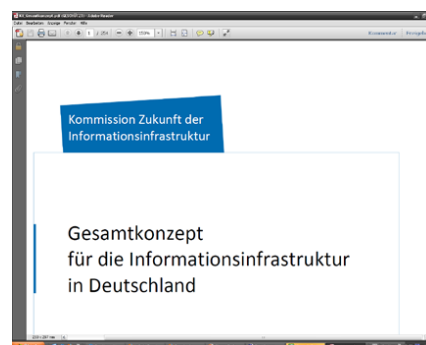
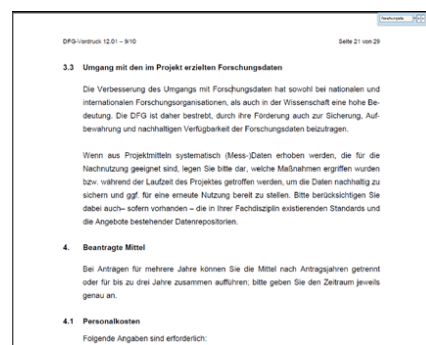
Einführung: Relevanz

➤ DFG (2010): Merkblatt und Leitfaden für Anträge

„Wenn aus Projektmitteln systematisch (Mess-)Daten erhoben werden, die für die Nachnutzung geeignet sind, legen Sie bitte dar, welche Maßnahmen ergriffen wurden bzw. während der Laufzeit des Projektes getroffen werden, um **die Daten nachhaltig zu sichern und ggf. für eine erneute Nutzung bereit zu stellen**. Bitte berücksichtigen Sie dabei auch – sofern vorhanden – die in Ihrer Fachdisziplin existierenden **Standards** und die Angebote **bestehender Datenrepositorien**.“

➤ KII-Papier - Gesamtkonzept zur Informationsinfrastruktur (2011): mit einem separaten Kapitel zum Thema **Forschungsdaten**

<http://www.wgl.de/?nid=infrastr&nidap=&print=0>





Einführung: Rückblick auf das Entstehungsumfeld (2007/2008)

Bibliothekartag Juni 2008 in Mannheim:

- „Bibliotheken und Forschungsdaten: Perspektiven“

Fazit-Folie (Auszug)

- **Erste positive Erfahrungen**
- **Chance für Bibliotheken:** Frühe Positionierung im Wissenschaftsprozess
- Mehrheitlich noch **kein Thema für Bibliotheken**
- **Herausforderungen**

Auszug aus dem Beitrag des Helmholtz Open Access Projekts (Pampel, Bertelmann, Hübner)

http://edoc.gfz-potsdam.de/gfz/get/11368/0/e870f73357d6c115e11ddccde37c0234/pampel_bibtag_080529_final.pdf



BW-eLabs: Wissensmanagement in virtuellen und remote Laboren

BW*eLabs*

- **Förderung:** Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg
- **Start:** September 2009, Laufzeit: 30 Monate
- **Ziel:** Erweiterung des Zugriffs auf heterogene, experimentelle Ressourcen, nachhaltige Erschließung und Nutzung von Forschungsdaten für Forschungs- bzw. Ausbildungszwecke



Projektpartner

BW*e*Labs

- Rechenzentrum der **Universität Stuttgart** (RUS)
 - Gesamtarchitektur, Projektleitung
- Institut für Technische Optik (ITO), **Universität Stuttgart**
 - Virtuelle Labore: *Digitale Holographie*
- Rechenzentrum und Freiburger Materialforschungszentrum (FMF) der **Universität Freiburg**
 - Bereitstellung der virtuellen und remote Labore: *Nanotechnologie*
- **FIZ Karlsruhe**
 - eSciDoc (Informations-, Kommunikations- und Publikationsplattform): Forschungsdatenrepository
- **Hochschule der Medien (HdM) Stuttgart**
 - Usability, Security

Projektziele (allgemein)

- Steigerung der Zugangsmöglichkeiten zu (nano-nahem) experimentellem Equipment
- Vernetzung und Integration verfügbarer virtueller und remote-kontrollierbarer Labore und Forschungsinformationen in einem kooperativen Wissensraum (Portal)
- Integriertes Dokumentenmanagement-System für die Archivierung und die Nachvollziehbarkeit von Forschungsdaten

BW*e*Labs





Motivation UB Stuttgart

Motivation: **Universität Stuttgart**

- Bündelung des Outputs von Wissenschaftlern der Universität Stuttgart
- Nachweis „aller“ Publikationen / Publikationsformen an einem Ort (Hochschulschriftenserver der Universität)
- Traditionelle Formen: Dissertationen, Habilitationen, Aufsätze ...
- ... aber nun zusätzlich auch: *Forschungsdaten(-nachweise)*

Motivation: **Bibliothekarische und Repository Community**

- OPUS-Community: Ausweitung der Funktionalitäten der OPUS-Software um eine Forschungsdaten-Komponente



Exkurs: OPUS

- OPUS ist eine **Open-Source-Software zum Betrieb von fachlichen und institutionellen Repositorien**
- Web-Anwendung, basierend auf Open-Source-Standardsoftware (PHP, MySQL, Apache, ...)
- Einfache Veröffentlichung, Erschließung, Administration und Recherche elektronischer Publikationen
- Offenes System: OAI- und weitere Schnittstellen
- **Credo** von Beginn an: „OPUS ist aufgrund seiner Architektur ein technisch **einfach zu installierendes und zu betreibendes System**, das im Produktionsbetrieb **sehr stabil und wartungsarm** läuft.“



BW-eLabs: UB Stuttgart

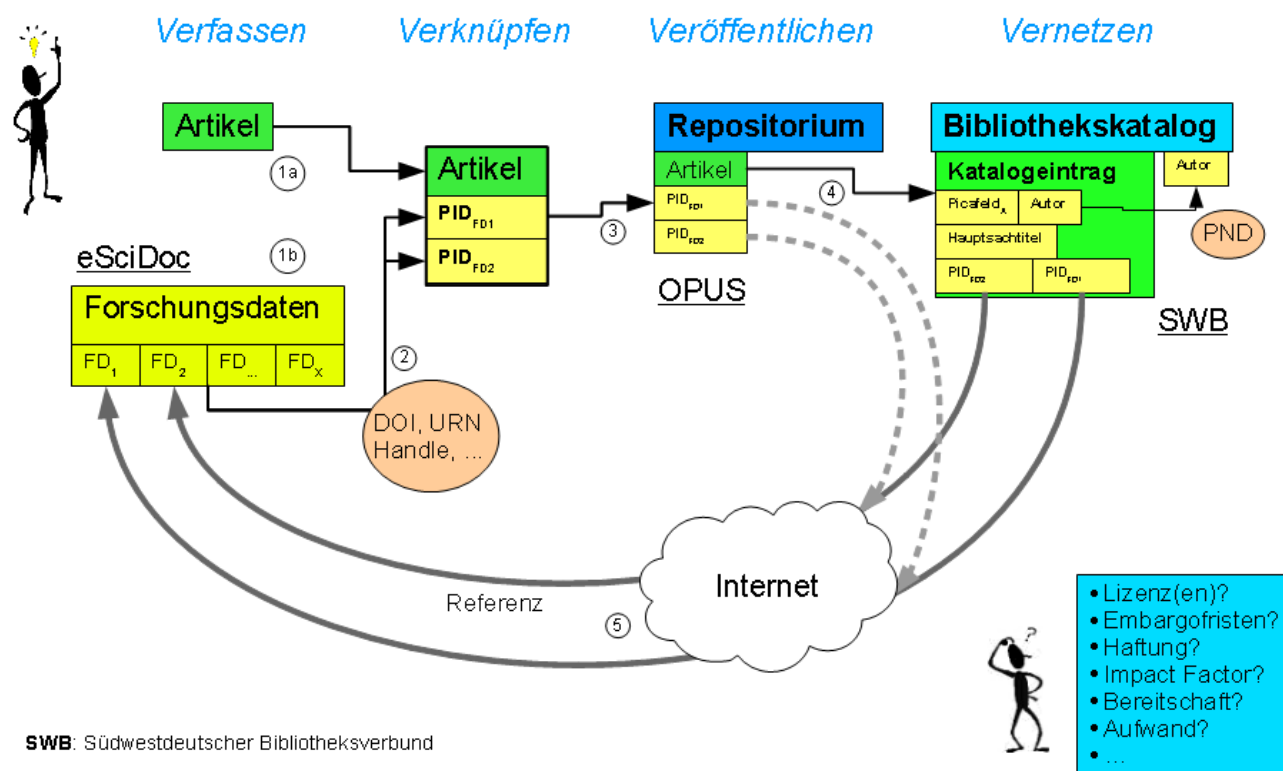
Aufgaben der UB Stuttgart:

- An- und Einbindung der Forschungsdaten an
 - OPUS, und damit ebenso in/an
 - Portale, Kataloge sowie
 - allgemein an die „bibliothekarische Community“

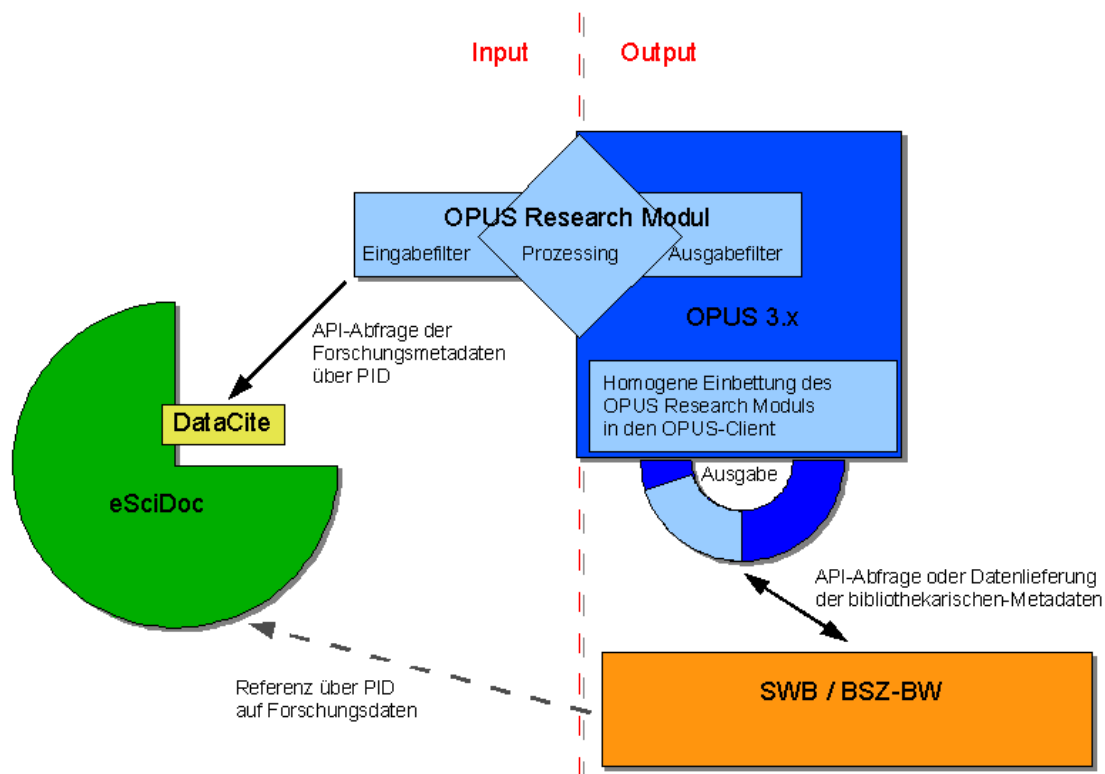
Konkrete Baustellen für die UB Stuttgart:

- Schnittstellen
- Metadaten-Mapping

BW-eLabs: Publikationsprozess



BW-eLabs / OPUS



Metadaten: FD – OPUS – SWB – DataCite (vereinfacht)

Nr.	Forschungsdaten [z.B.: FMF, ITO]	OPUS(-R)	SWB [BSZ] (Feldbezeichnung)	DataCite	Anmerkung
1	Name/Bezeichnung: Forschungsdatenobjekt	Titel	* Hauptsachtitel * Weitere Sachtitel	* Title	Name des Forschungsdatenobjekts
2	Beteiligte Personen	Person	*** 1.-3. Verfasser * sonstige Personen	* Creator / * Contributor	Autor(en), andere Beiträger; Bei DataCite sind hier auch Institutionen zugelassen; in den anderen Feldern nur Personen!
3	Institut(e), etc. der beteiligten Personen	Herausgeber	* Sekundärkörperschaft, Sonstige beteiligte Körperschaft	* Publisher / (* Contributor)	Nur Einrichtungen/Institute von beteiligten Personen (Autoren, etc.; an dieser Stelle werden keine Personen erfaßt! Außer bei DataCite.
4	Ursprungs-Institut und/oder -Repositorium	Institut (= Verlag)	* Erscheinungsort und Verlag	(* Publisher / * Creator)	Herkunftsnachweis (Entsprechung zu Verlag). Kann auch 3 entsprechen; bei DataCite auch Creator möglich
5	Fachgebiet	Wissenschaftliches Fachgebiet	* Schlagwörter aus einem Thesaurus (Fremddaten)	(* Subject)	Evt. Wikipedia-Klassen, in BW-eLabs noch nicht endgültig geklärt; BSZ: neues Feld geschaffen, noch nicht freigegeben!
6	Fachspezifische Klassifikation	Klassifikation (beliebig erweiterbar)	Derzeit nur in den Lokaldaten nicht im Verbund möglich	* Subject	Fachgebiets-, oder sonstige Klassifikationen wie z.B. PACS,
7	(Veröffentlichungs-) Datum	Datum (in jeder Form erfaßbar)	* Erscheinungsjahr	* Publication Year / (Date)	Hier wird das Publikationsdatum erfaßt. (Jahr/Monat/Tag); BSZ: Jahr ausreichend, komplettes Datum jedoch auch erfaßbar
8	DOI, Handle, URN, [weitere]	Jede Art von Persistent Identifier möglich	*URN, * DOI, * Handle; zugleich noch * Resolving URL	* Identifier (DOI) / * AlternateIdentifier	Jede Art von PID (Persistent Identifier); BSZ: Zugleich muß derzeit noch Resolving URL erfaßt werden, nötig für einige OPACs; DataCite: muß DOI sein
9	URL	URL	* Elektronische Adresse u. Zugriffsart für elektr. Ressource im Fernzugriff		URL, hier ist auch der Eintrag der URL-Auflösung des DOI möglich
10	(freie / vorgegebene) Schlagwörter	(alle) Key Words	* Schlagwortnormdatei * Schlagwörter aus einem Thesaurus (Fremddaten)	(* Subject)	Frei vergebene (auch englische) Schlagwörter; BSZ: SWD-Feld oder das neugeschaffene Feld zu nutzen, noch nicht freigegeben!
11	-	Datenträger	* Datenträger (cofi)	Evt. * ResourceType	Bei allen Objekten handelt es sich um „Computerdateien im Fernzugriff“, daher: cofi; Gilt für alle Objekte
12	-	Bibliographische Gattung	* Bibliographische Gattung / Status (oan)		Die Daten werden online als Datensatz von BW-eLabs an OPUS geliefert, daher: oan; Gilt für alle Objekte
13	Beschreibung, Abstract	Abstract, Summary	* Inhaltliche Zusammenfassung	* Description	Für Forschungsdaten extrem sinnvolles Feld. Wesentlicher Mehrwert für die Recherche, da im Katalog eine Volltextsuche möglich ist. Abstracts dieser Art sind verfügbar und kurz gehalten.

Fett bedeutet es handelt sich um ein Pflichtfeld * Bei SWB und DataCite steht jeweils für ein eigenes Feld **Besonders zu beachten beim SWB-Eintrag**

Schulze - ASpB-Tagung, Jülich, 9. November 2011



Ergebnisse und Nachhaltigkeit

Der Publikationsprozess über OPUS wird bis zum Projektende technisch gelöst sein:

- **Schnittstellen** zwischen den einzelnen Komponenten/Softwarelösungen: eSciDoc – OPUS – SWB-Verbundkatalog
- **Metadatenprofil(e)**: Absprachen/Lösungen zwischen UBS und FIZ, UBS und BSZ



Was haben wir erreicht? Was bleibt?

- Exemplarische Umsetzung
 - Erweiterung von OPUS (Research-Modul ...)
 - Schnittstellen
 - Workflows
- Pilotcharakter des Vorhabens: Identifikation von Themenfeldern und Herausforderungen
- Metadaten-Lösungen (GOLC, SWB)



GOLC: Metadaten für Remote Experimente und Forschungsdaten

GOLC Global Online Laboratory Consortium

Home Community News Contact Us Membership

Welcome!
The Global Online Laboratory Consortium

Home

Main Menu

- [Home](#)
- [Members](#)
- [Committees & Working Groups](#)
- [Events](#)
- [GOLC Wiki](#)
- [Related Links](#)
- [Technical Committee RFCs](#)
- [Register](#)

search...

Member Login

Username

Password

Remember Me ☐

[Login](#)

- [Forgot your password](#)
- [Forgot your username](#)
- [Create an account](#)

First GOLC Workshop: Improving Laboratory Learning Outcomes

The First Global Online Laboratory Workshop will be organized in conjunction with the [Frontiers in Education \(FIE\) 2011 Conference](#) in Rapid City, South Dakota, USA. Date: October 11 & 12, 2011

For more information >>[Click Here](#)

Or access the Workshop final program: [Workshop Program](#)

Registration for participation can be done here: <http://conftool.com/golc2011/>

Co-Sponsored by:

IEEE Education Society

The Global Online Laboratory Consortium is focused on promoting the development and sharing of, and research into remotely accessible laboratories for educational use. As the usage of online experiments gains traction in the educational community, there is increasing interest in developing online labs on a common infrastructure. A unified and interoperable architecture is essential to convert the current tremendous interest for online experiments into an economy of labs that can be efficiently shared around the world.

Mission Statement

"The mission of the consortium is the creation of sharable, online experimental environments which increase the educational and scientific value of learning which may not be accessible, scalable or efficient through traditional methods".

This means especially:

- to encourage and support the creation of new online labs and associated curricular materials;
- to sponsor the design of an efficient mechanism for sharing, exchanging and trading access to online labs by creation of a global network of shareable experiments;
- to support communities of scholars created around online laboratories; and

http://online-lab.org/index.php?option=com_content&view=article&id=35&Itemid=12



Herausforderungen

Herausforderungen bleiben, speziell hinsichtlich der

- organisatorischen und
 - rechtlichen Struktur zum Publizieren von Forschungsdaten
- = **Komplexität der Datenpublikation!**

Konkrete politisch-pragmatische Fragestellungen:

- Workflows für Abgabe und Publikation der Daten
- Infrastruktur für Persistent Identifier (DOI, DataCite)
- Wem gehören die Daten? Open Access auch Forschungsdaten?
- Rechtliches: Nutzungsrechte, Lizenzen ...
- Langzeitarchivierung
- Beratung für Wissenschaftler



Ausblick – Perspektiven (I): Forschungsdatenmanagement als Aufgabe von Bibliotheken

- BW-eLabs als „Leuchtturmprojekt“
- weitere Leuchtturmprojekte werden benötigt, gerade auch
 - im Themenfeld „Publikationsprozess - Forschungsdatenmanagement“
 - in unterschiedlichen, „datengetriebenen“ Disziplinen



Ausblick – Perspektiven (II): Forschungsdatenmanagement als Aufgabe von Bibliotheken

- Thema annehmen, bestehende (fachspezifische) **Angebote kennen - Beratungsfunktion**
- **Kooperationen** mit Rechenzentren, Verbund
- **Nachnutzung/Integration** bestehender Angebote: etwa Nachweis von Forschungsdaten im Katalog
- bei Bedarf:
 - **Konzepte** entwickeln **mit Wissenschaftlern** (Heterogenität der Fachkulturen)
 - Services für das Forschungsdatenmanagement (Erschließung und Nachweis, Qualitätssicherung, LZA, Qualifizierung ...)
 - **Forschungsdatenrepositorium**: ggf. Angebot entwickeln, falls noch keine (fachspezifischen) Lösungen für Forschungsdaten existieren



Science in the Open

The online home of Cameron Neylon

ABOUT BLOG PUBLICATIONS PRESENTATIONS CONTACT LIFESTREAM SPAM POLICY

Search the archive...

Home » Blog, Featured

Building the perfect data repository...or the one that might get used

27 OCTOBER 2011 1,134 VIEWS 5 COMMENTS

[A UV-vis readout of bis\(triphenylphosphine\) ni...](#)
 Image via Wikipedia

While there has been a lot of talk about data repositories and data publication there remains a real lack of good tools that are truly attractive to research scientists and also provide a route to more general and effective data sharing. [Peter Murray-Rust](#) has recently discussed the deficiencies of the traditional institutional repository as a research data repository in some depth [1, 2, 3, 4].

Data publication approaches and tools are appearing including [Dryad](#), [Figshare](#), [BuzzData](#) and more traditional venues such as [GigaScience](#) from [BioMedCentral](#) but these are all formal mechanisms that involve significant additional work alongside an active decision to "publish the data". The research repository of choice remains a haphazard file store and the data sharing mechanism of choice remains email. How do we bridge this gap?

One of the problems with many efforts in this space is how they are conceived and sold as the user. "Making it easy to put your data on the web" and "helping others to find your data" solve problems that most researchers don't think they have. Most researchers don't want to share at all, preferring to retain as much of an advantage through secrecy as possible. Those who do see a value in sharing are for the most part highly skeptical that the vast majority of research data can be used outside the lab in which it was generated. The small remainder who see a value in wider research data sharing are painfully aware of how much work it is to make that data useful.

A successful data repository system will start by solving a different problem, a problem that all researchers recognize they have, and will then nudge the users into doing the additional work of recording (or allowing the capture of) the metadata that could make that data useful to other researchers. Finally it will quietly encourage them to make the data accessible to other researchers. Both the nudge and the encouragement will arise by offering

License



To the extent possible under law, [Cameron Neylon](#) has waived all copyright and related or neighboring rights to Science in the Open. Published from the United Kingdom.

Recent Friendfeed activity

- Commented on Re: Building the perfect data repository... or the one that might get used Wednesday 6:06
- Shared: On thumb twiddling Tuesday 15:29
- Shared: (Some) garbage in, gold out Monday 16:23
- Shared: Open Access: a short summary Sunday 15:59
- Building the perfect data repository... or the one that might get used Thursday 21:45
- My FriendFeed profile ...

Lifestream

- RT @ricphillips: We send software and data into the future, not networks or hard drives. Nice thought from Rhys Francis at #eres2011 [cameronneylon] — 4h ago via Twitter
- Writing a blog post

I am also found at...

- Flickr
- Friendfeed
- GitHub
- My lab notebook
- Slideshare
- Twitter

<http://cameronneylon.net/blog/building-the-perfect-data-repository-or-the-one-that-might-get-used/>

Schulze - ASpB-Tagung, Jülich, 9. November 2011



- Vielen Dank für Ihre Aufmerksamkeit!
- Gibt es Fragen oder Kommentare?

Kontakt:

- Dr. Matthias Schulze,
Universitätsbibliothek Stuttgart
- matthias.schulze@ub.uni-stuttgart.de
- <http://www.ub.uni-stuttgart.de/wirueberuns/projekte>



http://www.carl-abrc.ca/about/working_groups/pdf/data_mgt_toolkit.pdf

Schulze - ASpB-Tagung, Jülich, 9. November 2011